SEQUENCE LISTING

- (1) GENERAL INFORMATION
- (i) APPLICANT: Lal, Preeti
 Bandman, Olga
- (ii) TITLE OF THE INVENTION: NOVEL HUMAN SODIUM-DEPENDENT PHOSPHATE COTRANSPORTER
- (iii) NUMBER OF SEQUENCES: 7
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
 - (B) STREET: 3174 Porter Drive
 - (C) CITY: Palo Alto
 - (D) STATE: CA
 - (E) COUNTRY: US
 - (F) ZIP: 94304
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ Version 2.0
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: To Be Assigned
 - (B) FILING DATE: Filed Herewith
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Billings, Lucy J.
 - (B) REGISTRATION NUMBER: 36,749
 - (C) REFERENCE/DOCKET NUMBER: PF-0221 US
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 415-855-0555
 - (B) TELEFAX: 415-845-4166
 - (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 401 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: BRAITUT02
 - (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEO ID NO:1:

Met Gln Val Asp Glu Thr Leu Ile Pro Arg Lys Val Pro Ser Leu Cys Ser Ala Arg Tyr Gly Ile Ala Leu Val Leu His Phe Cys Asn Phe Thr 25 Thr Ile Ala Gln Asn Val Ile Met Asn Ile Thr Met Val Ala Met Val 40 Asn Ser Thr Ser Pro Gln Ser Gln Leu Asn Asp Ser Ser Glu Val Leu Pro Val Asp Ser Phe Gly Gly Leu Ser Lys Ala Pro Lys Ser Leu Pro 70 75 Ala Lys Ser Ser Ile Leu Gly Gly Gln Phe Ala Ile Trp Glu Arg Trp 90 Gly Pro Pro Gln Glu Arg Ser Arg Leu Cys Ser Ile Ala Leu Ser Gly 105 Met Leu Leu Gly Cys Phe Thr Ala Ile Leu Ile Gly Gly Phe Ile Ser 120 Glu Thr Leu Gly Trp Pro Phe Val Phe Tyr Ile Phe Gly Gly Val Gly 135 Cys Val Cys Cys Leu Leu Trp Phe Val Val Ile Tyr Asp Asp Pro Val 150 155 Ser Tyr Pro Trp Ile Ser Thr Ser Glu Lys Glu Tyr Ile Ile Ser Ser 165 170 Leu Lys Gln Gln Val Gly Ser Ser Lys Gln Pro Leu Pro Ile Lys Ala 180 185 Met Leu Arg Ser Leu Pro Ile Trp Ser Ile Cys Leu Gly Cys Phe Ser 200 205 His Gln Trp Leu Val Ser Thr Met Val Val Tyr Ile Pro Thr Tyr Ile 215 220 Ser Ser Val Tyr His Val Asn Ile Arg Asp Asn Gly Leu Leu Ser Ala 230 235 Leu Pro Phe Ile Val Ala Trp Val Ile Gly Met Val Gly Gly Tyr Leu 245 250 Ala Asp Phe Leu Leu Thr Lys Lys Phe Arg Leu Ile Thr Val Arg Lys 265 Ile Ala Thr Ile Leu Gly Ser Leu Pro Ser Ser Ala Leu Ile Val Ser 280 285 Leu Pro Tyr Leu Asn Ser Gly Tyr Ile Thr Ala Thr Ala Leu Leu Thr 295 300 Leu Ser Cys Gly Leu Ser Thr Leu Cys Gln Ser Gly Ile Tyr Ile Asn 310 315 Val Leu Asp Ile Ala Pro Arg Tyr Ser Ser Phe Leu Met Gly Ala Ser 325 330 Arg Gly Phe Ser Ser Ile Ala Pro Val Ile Val Pro Thr Val Ser Gly 340 345 Phe Leu Leu Ser Gln Asp Pro Glu Phe Gly Trp Arg Asn Val Phe Phe 360 Leu Leu Phe Ala Val Asn Leu Leu Gly Leu Leu Phe Tyr Leu Ile Phe 375 380 Gly Glu Ala Asp Val Gln Glu Trp Ala Lys Glu Arg Lys Leu Thr Arg 395 Leu

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1643 base pairs



- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT02
- (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTCCGGCAAT TCCCGGGTCG ACCCACGCGT CCGCTTGGAG GGACGCTGGG TTCAACTTGA AGCCCTTCCA CAGACATTAA GTCGGTGAAA ACCATTCACT AGGAGAGGAG	AGAACGGTGA	GGATGACCGA	CGTATAGGCG	AGAGCCTAGG	TACGCCATGC	CAGGTCACCG	60
GCCACCAGA CAGAGTTGAG TCCCACAGCA AGGGAGAGCA AGAACGCACA AGATATGCAA GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAAGTTTAT GTTCTGCTCG CTATGGAATA 300 GCCCTCGTCT TACATTTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG GTGCTGCCTG TTGACTCAT TGGTGGCCTA AGTAAAGCCC CAAAGAGGTCT TCCTTGCAAGA TCCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG GTTTTTAAGGC AGCCCTTCC CATCAAAGCT ATGCTCAGAT CTTTACCAGC TTTACCAGCA TCTTTACAGCA TCTTTACAGCA TCTTTACAGCA TCTTTACAGCA TCATTAGCACA TGGTTGATAT CTTTGGAGGT 780 TCTTTTAGGCT TCTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTTGACCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TCTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCCATTAT CCTTCTAACCC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTCCCACAA TTTTAGGAAG TCCCCTCT 1080 TCAGCACTCA TTGTGTCCT GCCTTACCTC AATTCCGCT ATTTAGGAAG TCCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGCT ATATCCACC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGCACA TTTTAGGAAG TCCCCCTCT 1080 GATATTGCC CAAGGTATC CAGTTTCCC AATTCCTCAAGCA TGCAGAGATT TCCACCATT TCCACCAT TTGAGCACA TGTCAGCACA TTTTAGGAACA TCCCCCTCT 1080 GATATTGCC CAAGGATATC CAGTTTTCTC ATGGGACACT CAAGAGGATT TCCACCCTC 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGCACA TTTTTAGAGAC TCTCCCCTCT 1080 GATATTGCC CAAGGATATC CAGTTTTCC ATGGGACACT CAAGAGGATT TCCACCTC 1380 CCGTGATGAA AAGCACATTG TGTCAGCGGA TTTCTTCTTA GTCAGCAC TGTTTTTGG AACCCTTCTTTTTCTT GCTGTTTTCC TTTCTCTT TAGGACACA TTGCATCAC TCTTCTACCTC 1380 CCGTGATGAA AAGCACTTC TTTCTTCT TAGGCACAT TTGCATCACA TCTTCTACCTC 1380 CCGTGATGAA AAGCAGATG CAAAGATCTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAA AATACCAGTG AAAAGATCTT TTTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560	GTCCGGCAAT	TCCCGGGTCG	ACCCACGCGT	CCGCTTGGAG	GGACGCTGGG	TTCAACTTGA	120
GTGGATGAGA CACTGATCCC CAGGAAAGTT CCAAGTTTAT GTTCTGCTCG CTATGGAATA 300 GCCCTCGTCT TACATTTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACATC GCAGCACTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCT GCATGGTT TCTTGTGAG ATGTTACTGG GATGCTTTAC CTTTGGAGGT 660 ATAGGTGGT TCTTGTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGT TCTTCTGAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTAC AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTAC AGAAGAACAA TGGTTGATAA CATACCAACT 900 TACATCAGCT CTGTGACCA TCAATGGTTA GTTAGCACAA TGGTTGATAA CATACCAACT 900 TACATCAGCT CTGTGACCA TCAATGGTTA GTTAGCACAA TGGTTGATAA CATACCAACT 900 TACATCAGCT CTGTGACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 1080 TAAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCCA AATTCCCGCT ATATCACAGC AACTGCCTTG 1140 CTGACGACTCA TTGTGTCTCT GCCTTTCCC AATTCCCGCT ATATCACAGC AACTGCCTTG 1140 CTGACGACTCA TTGTGCCATTC CAGTTTTCTC AATTCCCGC AATTCACAGC AACTGCCTTG 1140 GAACACTCCA TTGTACCCA TGTGAGGAAA ATTGCCACAA TTTTAGGAAG AACTGCCTTG 1140 GAACACTCCA TTGTACCCA TGTGAGGGAA ATTGCCACAA TTTTAGGAAC CAAGAGTAT TCCCCCTCT TAGGACACA TTGTACCCA TGTGAGGAAA ATTGCCACA TTATATAT CAATGTCTTA 1200 GATATTGCC CAAGGTATT CAAGAGAATG GACTTCTTCTT GCCTTTCTC TGCCTTCTCC TGTGAGGAAA ATTGCCACA TTTTAGGAAC CTGACTTTGGG 1320 TGGAGGATAT TCTTCTTCT GCCTTTCCC TGTTTGCC TTTACCCA TGACTTTGGG 1320 TGGAGGATAT TCTTCTTCT GCCTTTCTCC TGTTTGCC TTTAACCAC TGACTTTAGGA 1440 AGTTATCCCA CCTTGGAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGAGAGA AAAAGTCATT TTTTTCCTGT TGGCCTTTTC AATAAAAATTCA 1560 CCGTGATTAAAAA TAAAAAATTCA 1560 CCGTGATTAACCAT TAAAAAATTCA TTTTTTTTTTTT	AGCCCTTCCA	CAGACATTAA	GTCGGTGAAA	ACCATTCACT	AGGAGAGGAG	AAACACAATG	180
GCCCTCGTCT TACATTCTG CAATTTCACA ACGATAGCAC AAAATGTCAT CATGAACATC 360 ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCT 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTTGTGCACA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAAGC AGCCTCTCC CATCAAAGCA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCAC TCATAACAC ACAGGTCGGG 780 TACATCAGCT CTGTGTACCA TGTTAACACA AGAGACAAC ACAGGTCGGG 780 TACATCAGCT CTGTGTACCA TGTTAACACA AGAGACAAC ACAGGTCGGG 780 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAAATG GACTTCTATC TGCCCTTCCT 960 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TACATCAGCT CTGTGTACCA TGTGAGGAAAAAGGTTTA AGGCACTCA TTGTGTCCATA AGGCATGGTG GAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCT AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTC TTGGGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGCCTTA 1200 GATATTGCC CAAGGTATT CAGTTTTCTC ATGTCAGCAC TGTCAGCGAC TTTCTTCTA CTGCGCATA TCTGAGCACA TTTTTTTCT GCCCTTCT TTTTTTTCT GCCCTTCT TAGGAACA TTCTCAGCCC TGAGTTTTGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTCTC GTTAACCAC TGTCAGCAC TTTCTTCTA TTCGAGACCT TTTCTACCTC 1380 ATATTTGGAA AAAAGATTTT TTTTTCTTTT TGAGAAAAT TAAAAATTCA 1500 TCCCTTTTTT TTTTTTCTTTT TGAGAGAAAA TAAAAATTCA 1500 TCCCTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTT	GCCACCAAGA	CAGAGTTGAG	TCCCACAGCA	AGGGAGAGCA	AGAACGCACA	AGATATGCAA	240
ACCATGGTAG CCATGGTCAA CAGCACAAGC CCTCAATCCC AGCTCAATGA TTCCTCTGAG 420 GTGCTGCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTTGTGTACCA TGTTAACATC AGAGACAAAG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGT GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCACCA TGTGAGGAAA ATTGCCACAA TTTTAGGAGA TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCCTC GCCTTACCTC AATTCCGGCT ATATCACAC TGTGAGGAAA ATTGCCACAA TTTTAGCAGA AACTGCCTTC 1080 GATATTGCTC CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGCCTT 1140 GCACCTCA TTGTGCGGATT CAGCACATTG TGTCAGTCAG GGATTTATAT CAATGCCTT 1140 GCACCTCC CAAGGTATTC CAGCTTTCTC TGCCCTTCT TGCCCTTCT 1080 GATATTGCCC CAAGGTATTC CAGCTTTCTC AATTCCGGC TGAGTTTAGG 1320 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTCGAGCATA TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCAGA CAGCGCTTT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCAGA CAGCGCTT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTCTCACCTC TAGGAGCAC TGAGAGTATT TCCAGCACA TTTCTACCTC TAGGAGCAC TGAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCTC AAGAGAGA GAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGA AAAAGATTT TTTTTCCTGT GGCTCTTTTC AATTAGAGA TAAAAATTCA 1500 CCGTGATGAA AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTGAA AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT AGAAAGATTT TTTTTCCTGT GGCTCTTTTC AATTATGAA 1560 TCAGGTTCATT AGAAAGATTT TTTTTCCTGT GGCTCTTTTC AATTATGAA 1560 TCAGGTTCATT AGAAAATTCA TTAAAAAATTCA 1560 TCAGGTTCATT TTTTTTTTTTTTTTTTTTTTTTTTTTT	GTGGATGAGA	CACTGATCCC	CAGGAAAGTT	CCAAGTTTAT	${\tt GTTCTGCTCG}$	CTATGGAATA	300
GTGCTGCCTG TTGACTCATT TGGTGGCCTA AGTAAAGCCC CAAAGAGTCT TCCTGCAAAG 480 TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTCC ACAAGAACGA 540 AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAAAG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGT GGAGGCTATC TGGCAGATT CCTTCTCAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAAGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCCT GCCTTACCT AATTCCGGCT ATTTAGGAAG TCTCCCCTCT 1080 GAAAAAGTTTA GACTCATCCC TGTGAGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 GAAAAAGTTC CAAGGATTC CAAGCACATTG TGCAGCACA TTTTAGGAAG TCTCCCCTCT 1080 GAAAAAGTTC CAAGGATTC CAAGCACATTG TGCAGCACA TTTCAGCAC AACTGCCTTG 1140 GAAAAGTTCC CAAGGATTC CAGGTATTC CAGGTATTATAT CAATGCTTA 1200 GATATTGCCC CAAGGAATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTGCAGCATA TTCCAGCCATA 1260 GCACCTGTCA TTGTACCCC TGTCAGCGGA TTTCTTCTTA GTCAGCACA TTTCTACCTC TAGGAGCATC TTCCAGCCGT TAGGAGCATC TTCTACCTC 1380 ATATTTGGAG AAGCACATTT TCCAGCGGA TTTCTTCTTA GTCAGCACA TTCTTATGA 1440 AGTTATCCCA CCTTGGATGA AAAAGATTT TTTTTCCTGT GGCTCTTTC AATTAGAGA TAGAAAGTCT 1500 CCGTGATGAA AATACCAGTG AAAAGATTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAA AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAA AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAAAAATCAC TCGTTTATGA 1440 AGTTTTCTTTTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 CCGTGATGAAAAATCAC TCGTTTATGA 1560 CCGTGATGAAAAATCAC TCGTTTATGA 1560 CCGTGATGAAAAATCAC TCGTTTATGA 1560 CCGTGATGAAAAAATTCA TAAAAAATTCA 1560 CCGTGATTCATT TTTTTTCCTGT GGCTCTTTTC AATTA	GCCCTCGTCT	TACATTTCTG	CAATTTCACA	ACGATAGCAC	AAAATGTCAT	CATGAACATC	360
TCCTCAATAC TTGGGGGTCA GTTTGCAATT TGGGAAAGGT GGGGCCCTC ACAAGAACGA AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC 600 ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTG TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAAT GTAAGATGAA TAAAAATTCA 1620	ACCATGGTAG	CCATGGTCAA	CAGCACAAGC	CCTCAATCCC	AGCTCAATGA	TTCCTCTGAG	420
AGCAGACTCT GCAGCATTGC TTTATCAGGA ATGTTACTGG GATGCTTTAC TGCCATCCTC ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGCTACTC ATGCAGCACA TTCTCAGCC TGAGTTTGGG 1320 TGGAGGAATG TCTCCCTCT TAGGAGGAATG TCTCCCTCT TAGGAGGAATG TCTCTTACCTC ATGGAGGAAC CAAGAGGATT TCCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTACCT TAGGAGGACC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTTCC GTTAACCTC TAGGACCAC TCTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACCTTTTTT TTTTTCCTGT GGCTCTTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACCTTTTTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACCTTTTTTT TTTTTCCTGT GGCTCTTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTTTTTTT	GTGCTGCCTG	TTGACTCATT	TGGTGGCCTA	AGTAAAGCCC	CAAAGAGTCT	TCCTGCAAAG	480
ATAGGTGGCT TCATTAGTGA AACCCTTGGG TGGCCCTTTG TCTTCTATAT CTTTGGAGGT 660 GTTGGCTGTG TCTGCTGCCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTGGAG AAGCAGATGT CCAAGAATAG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	TCCTCAATAC	TTGGGGGTCA	GTTTGCAATT	${\tt TGGGAAAGGT}$	GGGGCCCTCC	ACAAGAACGA	540
GTTGGCTGTG TCTGCTGCT TCTCTGGTTT GTTGTGATTT ATGATGACCC CGTTTCCTAT 720 CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGATCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACCTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	AGCAGACTCT	GCAGCATTGC	TTTATCAGGA	${\tt ATGTTACTGG}$	GATGCTTTAC	TGCCATCCTC	600
CCATGGATAA GCACCTCAGA AAAAGAATAC ATCATATCCT CCTTGAAACA ACAGGTCGGG 780 TCTTCTAAGC AGCCTCTTCC CATCAAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA 840 TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGGAGAAAT GTAAGAATGAA TAAAAAATTCA 1620	ATAGGTGGCT	TCATTAGTGA	AACCCTTGGG	${\tt TGGCCCTTTG}$	TCTTCTATAT	CTTTGGAGGT	660
TCTTCTAAGC AGCCTCTTCC CATCAAGCT ATGCTCAGAT CTCTACCCAT TTGGTCCATA TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GTTGGCTGTG	TCTGCTGCCT	TCTCTGGTTT	${\tt GTTGTGATTT}$	ATGATGACCC	CGTTTCCTAT	720
TGTTTAGGCT GTTTCAGCCA TCAATGGTTA GTTAGCACAA TGGTTGTATA CATACCAACT 900 TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT 960 TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	CCATGGATAA	GCACCTCAGA	AAAAGAATAC	ATCATATCCT	CCTTGAAACA	ACAGGTCGGG	780
TACATCAGCT CTGTGTACCA TGTTAACATC AGAGACAATG GACTTCTATC TGCCCTTCCT TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC 1020 AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TCTTCTAAGC	AGCCTCTTCC	CATCAAAGCT	ATGCTCAGAT	CTCTACCCAT	TTGGTCCATA	840
TTTATTGTTG CCTGGGTCAT AGGCATGGTG GGAGGCTATC TGGCAGATTT CCTTCTAACC AAAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TGTTTAGGCT	GTTTCAGCCA	TCAATGGTTA	GTTAGCACAA	${\tt TGGTTGTATA}$	CATACCAACT	900
AAAAAGTTTA GACTCATCAC TGTGAGGAAA ATTGCCACAA TTTTAGGAAG TCTCCCCTCT 1080 TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TACATCAGCT	CTGTGTACCA	TGTTAACATC	AGAGACAATG	GACTTCTATC	TGCCCTTCCT	960
TCAGCACTCA TTGTGTCTCT GCCTTACCTC AATTCCGGCT ATATCACAGC AACTGCCTTG 1140 CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAAT GTAAGATGAA TAAAAAATTCA 1620	TTTATTGTTG	CCTGGGTCAT	AGGCATGGTG	GGAGGCTATC	${\tt TGGCAGATTT}$	CCTTCTAACC	1020
CTGACGCTCT CTTGCGGATT AAGCACATTG TGTCAGTCAG GGATTTATAT CAATGTCTTA 1200 GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	AAAAAGTTTA	GACTCATCAC	TGTGAGGAAA	ATTGCCACAA	${\tt TTTTAGGAAG}$	TCTCCCCTCT	1080
GATATTGCTC CAAGGTATTC CAGTTTTCTC ATGGGAGCAT CAAGAGGATT TTCGAGCATA 1260 GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TCAGCACTCA	TTGTGTCTCT	GCCTTACCTC	AATTCCGGCT	ATATCACAGC	AACTGCCTTG	1140
GCACCTGTCA TTGTACCCAC TGTCAGCGGA TTTCTTCTTA GTCAGGACCC TGAGTTTGGG 1320 TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	CTGACGCTCT	CTTGCGGATT	AAGCACATTG	TGTCAGTCAG	${\tt GGATTTATAT}$	CAATGTCTTA	1200
TGGAGGAATG TCTTCTTCTT GCTGTTTGCC GTTAACCTGT TAGGACTACT CTTCTACCTC 1380 ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GATATTGCTC	CAAGGTATTC	CAGTTTTCTC	ATGGGAGCAT	CAAGAGGATT	TTCGAGCATA	1260
ATATTTGGAG AAGCAGATGT CCAAGAATGG GCTAAAGAGA GAAAACTCAC TCGTTTATGA 1440 AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	GCACCTGTCA	TTGTACCCAC	TGTCAGCGGA	${\tt TTTCTTCTTA}$	GTCAGGACCC	TGAGTTTGGG	1320
AGTTATCCCA CCTTGGATGG AAAAGTCATT AGGCACCGTA TTGCATAAAA TAGAAGGCTT 1500 CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	TGGAGGAATG	TCTTCTTCTT	GCTGTTTGCC	GTTAACCTGT	TAGGACTACT	CTTCTACCTC	1380
CCGTGATGAA AATACCAGTG AAAAGATTTT TTTTTCCTGT GGCTCTTTTC AATTATGAGA 1560 TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAAATTCA 1620	ATATTTGGAG	AAGCAGATGT	CCAAGAATGG	GCTAAAGAGA	GAAAACTCAC	TCGTTTATGA	1440
TCAGTTCATT ATTTTATTCA GACTTTTTT TGAGAGAAAT GTAAGATGAA TAAAAATTCA 1620	AGTTATCCCA	CCTTGGATGG	AAAAGTCATT	AGGCACCGTA	TTGCATAAAA	TAGAAGGCTT	1500
	CCGTGATGAA	AATACCAGTG	AAAAGATTTT	${\tt TTTTTTCCTGT}$	GGCTCTTTTC	AATTATGAGA	1560
	TCAGTTCATT	ATTTTATTCA	GACTTTTTTT	TGAGAGAAAT	${\tt GTAAGATGAA}$	TAAAAATTCA	1620
AATAAAATGA TAACTAAGAA TGC 1643	AATAAAATGA	TAACTAAGAA	TGC				1643

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 467 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (A) LIBRARY: GenBank
 - (B) CLONE: 450532
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

 Met Gln Met Asp Asn Arg Leu Pro Pro Lys Lys Val Pro Gly Phe Cys

 1
 5
 10
 15

 Ser Phe Arg Tyr Gly Leu Ser Phe Leu Val His Cys Cys Asn Val Ile
 20
 25
 30

 Ile Thr Ala Gln Arg Ala Cys Leu Asn Leu Thr Met Val Val Met Val
 35
 40
 45



Asn Ser Thr Asp Pro His Gly Leu Pro Asn Thr Ser Thr Lys Lys Leu 55 60 Leu Asp Asn Ile Lys Asn Pro Met Tyr Asn Trp Ser Pro Asp Ile Gln 70 75 Gly Ile Ile Leu Ser Ser Thr Ser Tyr Gly Val Ile Ile Ile Gln Val 90 Pro Val Gly Tyr Phe Ser Gly Ile Tyr Ser Thr Lys Lys Met Ile Gly 100 105 Phe Ala Leu Cys Leu Ser Ser Val Leu Ser Leu Ile Pro Pro Ala 120 Ala Gly Ile Gly Val Ala Trp Val Val Cys Arg Ala Val Gln Gly 135 Ala Ala Gln Gly Ile Val Ala Thr Ala Gln Phe Glu Ile Tyr Val Lys 150 155 Trp Ala Pro Pro Leu Glu Arg Gly Arg Leu Thr Ser Met Ser Thr Ser 170 Gly Phe Leu Leu Gly Pro Phe Ile Val Leu Leu Val Thr Gly Val Ile 180 185 190 Cys Glu Ser Leu Gly Trp Pro Met Val Phe Tyr Ile Phe Gly Ala Cys 200 Gly Cys Ala Val Cys Leu Leu Trp Phe Val Leu Phe Tyr Asp Asp Pro 215 Lys Asp His Pro Cys Ile Ser Ile Ser Glu Lys Glu Tyr Ile Thr Ser 230 235 Ser Leu Val Gln Gln Val Ser Ser Ser Arg Gln Ser Leu Pro Ile Lys 245 250 Ala Ile Leu Lys Ser Leu Pro Val Trp Ala Ile Ser Ile Gly Ser Phe 260 265 Thr Phe Phe Trp Ser His Asn Ile Met Thr Leu Tyr Thr Pro Met Phe 280 Ile Asn Ser Met Leu His Val Asn Ile Lys Glu Asn Gly Phe Leu Ser 295 300 Ser Leu Pro Tyr Leu Phe Ala Trp Ile Cys Gly Asn Leu Ala Gly Gln 310 315 Leu Ser Asp Phe Phe Leu Thr Arg Asn Ile Leu Ser Val Ile Ala Val 330 325 Arg Lys Leu Phe Thr Ala Ala Gly Phe Leu Leu Pro Ala Ile Phe Gly 345 Val Cys Leu Pro Tyr Leu Ser Ser Thr Phe Tyr Ser Ile Val Ile Phe 360 365 Leu Ile Leu Ala Gly Ala Thr Gly Ser Phe Cys Leu Gly Gly Val Phe 375 Ile Asn Gly Leu Asp Ile Ala Pro Arg Tyr Phe Gly Phe Ile Lys Ala 390 395 Cys Ser Thr Leu Thr Gly Met Ile Gly Gly Leu Ile Ala Ser Thr Leu 410 405 Thr Gly Leu Ile Leu Lys Gln Asp Pro Glu Ser Ala Trp Phe Lys Thr 420 425 Phe Ile Leu Met Ala Ala Ile Asn Val Thr Gly Leu Ile Phe Tyr Leu 440 Ile Val Ala Thr Ala Glu Ile Gln Asp Trp Ala Lys Glu Lys Gln His 455 460 Thr Arg Leu 465

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 560 amino acids





- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 507415

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Glu Phe Arg Gln Glu Glu Phe Arg Lys Leu Ala Gly Arg Ala Leu Gly Arg Leu His Arg Leu Leu Glu Lys Arg Gln Glu Gly Ala Glu Thr 20 25 Leu Glu Leu Ser Ala Asp Gly Arg Pro Val Thr Thr His Thr Arg Asp 40 Pro Pro Val Val Asp Cys Thr Cys Phe Gly Leu Pro Arg Arg Tyr Ile Ile Ala Ile Met Ser Gly Leu Gly Phe Cys Ile Ser Phe Gly Ile Arg 70 75 Cys Asn Leu Gly Val Ala Ile Val Ser Met Val Asn Asn Ser Thr Thr 90 His Arg Gly Gly His Val Val Gln Lys Ala Gln Phe Asn Trp Asp 100 105 Pro Glu Thr Val Gly Leu Ile His Gly Ser Phe Phe Trp Gly Tyr Ile 120 125 Val Thr Gln Ile Pro Gly Gly Phe Ile Cys Gln Lys Phe Ala Ala Asn 135 140 Arg Val Phe Gly Phe Ala Ile Val Ala Thr Ser Thr Leu Asn Met Leu 150 155 Ile Pro Ser Ala Ala Arg Val His Tyr Gly Cys Val Ile Phe Val Arg 165 170 Ile Leu Gln Gly Leu Val Glu Gly Val Thr Tyr Pro Ala Cys His Gly 180 185 Ile Trp Ser Lys Trp Ala Pro Pro Leu Glu Arg Ser Arg Leu Ala Thr 200 205 Thr Ala Phe Cys Gly Ser Tyr Ala Gly Ala Val Val Ala Met Pro Leu 215 220 Ala Gly Val Leu Val Gln Tyr Ser Gly Trp Ser Ser Val Phe Tyr Val 230 235 Tyr Gly Ser Phe Gly Ile Phe Trp Tyr Leu Phe Trp Leu Leu Val Ser 245 250 Tyr Glu Ser Pro Ala Leu His Pro Ser Ile Ser Glu Glu Glu Arg Lys 260 265 Tyr Ile Glu Asp Ala Ile Gly Glu Ser Ala Lys Leu Met Asn Pro Val 280 Thr Lys Phe Asn Thr Pro Trp Arg Arg Phe Phe Thr Ser Met Pro Val 295 290 300 Tyr Ala Ile Ile Val Ala Asn Phe Cys Arg Ser Trp Thr Phe Tyr Leu 310 315 Leu Leu Ile Ser Gln Pro Ala Tyr Phe Glu Glu Val Phe Gly Phe Glu 325 330 Ile Ser Lys Val Gly Leu Val Ser Ala Leu Pro His Leu Val Met Thr 345 Ile Ile Val Pro Ile Gly Gly Gln Ile Ala Asp Phe Leu Arg Ser Arg 360 His Ile Met Ser Thr Thr Asn Val Arg Lys Leu Met Asn Cys Gly Gly 375 380 Phe Gly Met Glu Ala Thr Leu Leu Val Val Gly Tyr Ser His Ser





385					390					395					400
Lys	Gly	Val	Ala	Ile 405	Ser	Phe	Leu	Val	Leu 410	Ala	Val	Gly	Phe	Ser 415	Gly
Phe	Ala	Ile	Ser 420		Phe	Asn	Val	Asn 425		Leu	Asp	Ile	Ala 430		Arg
Tyr	Ala	Ser 435		Leu	Met	Gly	Ile 440	Ser	Asn	Gly	Val	Gly 445		Leu	Ser
Gly	Met 450	Val	Cys	Pro	Ile	Ile 455	Val	Gly	Ala	Met	Thr 460	Lys	His	Lys	Thr
Arg 465	Glu	Glu	Trp	Gln	Tyr 470	Val	Phe	Leu	Ile	Ala 475	Ser	Leu	Val	His	Tyr 480
Gly	Gly	Val	Ile	Phe 485	Tyr	Gly	Val	Phe	Ala 490	Ser	Gly	Glu	Lys	Gln 495	Pro
Trp	Ala	Glu	Pro 500	Glu	Glu	Met	Ser	Glu 505	Glu	Lys	Сув	Gly	Phe 510	Val	Gly
His	Asp	Gln 515	Leu	Ala	Gly	Ser	Asp 520	Glu	Ser	Glu	Met	Glu 525	Asp	Glu	Val
Glu	Pro 530	Pro	Gly	Ala	Pro	Pro 535	Ala	Pro	Pro	Pro	Ser 540	Tyr	Gly	Ala	Thr
His 545	Ser	Thr	Val	Gln	Pro 550	Pro	Arg	Pro	Pro	Pro 555	Pro	Val	Arg	Asp	Tyr 560

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 272 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: BRAITUT02
- (B) CLONE: 754412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTTATATCA	ATGTCTTAGA	TATTGCTCCA	AGGTATTCCA	GTTTTCTCAT	GGGAGCATCA	60
AGAGGATTTT	CGAGCATAGC	ACCTGTCATT	GTACCCACTG	TCAGTGGATT	TCTTCTTAGT	120
CAGGACCCTG	AGTTTGGGTG	GAGGAATGTC	${\tt TTCTTCTTGC}$	TGTTTGCCGT	TAACCTGTTA	180
GGACTACTCT	TCTACCTCAT	ATTTGGAGAA	GCAGATGTCC	AAGAATGGGC	TAAAGAGAGA	240
AAACTCACTC	GTTTATGAAG	TTATCCCACC	TT			272

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: XLR
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTTGATGCTC CCATGAGAAA ACTGG

(2) INFORMATION FOR SEQ ID NO:7:

25





- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: XLF
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGGATTTTCG AGCATAGCAC CTGTC

25